Accelerating MCMC algorithms by breaking detailed balance

Konstantin Turitsyn

T-4 & CNLS, Los Alamos National Laboratory
Landau Institute for Theoretical Physics, Moscow

In collaboration with

Misha Chertkov (LANL)
Marija Vucelja (Weizmann)

http://arXiv.org/abs/0809.0916

MCMC Algorithms

- Problem: produce samples x from a given distribution π_x defined up to a constant.
- Solution: use a Markovian random walk x^t converging to the required stationary distribution.
- Detailed balance: $T_{yx}\,\pi_x = T_{xy}\,\pi_y$
- Only local moves are allowed: x^{t+1} is close to x^t

Detailed balance

Random walker distribution evolves according to

$$P_x^{t+1} = \sum_y T_{xy} P_y^t$$

Assuming $\sum_x T_{xy}$ 1 one can rewrite the equation as

$$\sum_y \left[T_{yx} P_x^{t+1} - T_{xy} P_y^t \right] = 0$$

Balance condition: $\forall t: \; P_x^t = \pi_x$

$$\sum_y \left[T_{yx} \pi_x - T_{xy} \pi_y \right] = 0$$

Detailed balance (reversibility, equilibrium): $T_{yx} \pi_x = T_{xy} \pi_y$

Sufficient but not necessary!

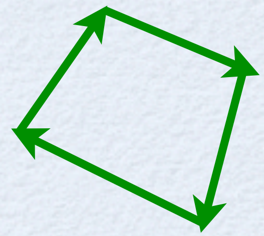
Loop decomposition

It is useful to consider ergodic flux matrix $Q_{xy} = T_{xy}\pi_y$

Detailed balance = symmetry of ergodic flow: $Q_{xy} = Q_{yx}$

Asymmetric part of Q_{xy} can be decomposed:

$$Q_{xy} - Q_{yx} = \sum_{\alpha} J_\alpha \left(C^\alpha_{xy} - C^\alpha_{yx}\right)$$



Here J_α is amplitude of the probability flow

C^α_{xy} is adjacency matrix of the loop

Flow amplitudes are bounded by the reversible part

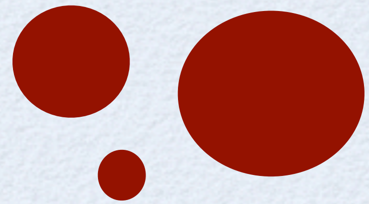
Physical analogies

- PDF evolution ⇔ Diffusion-advection of passive scalar
- Balance condition ⇔ Flow Incompressibility
- Reversibility ⇔ Diffusion
- Irreversible motion ⇔ Advection
- Loops ⇔ Vortices

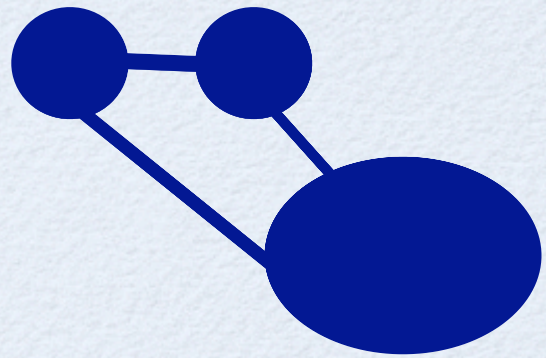
Slow convergence

Several types of distributions are characterized by slow mixing:

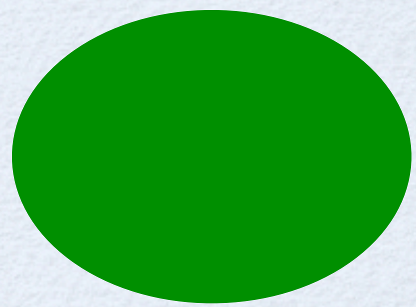
- Glassy landscapes: Regions that dominate the partition function are separated by “energy barriers”



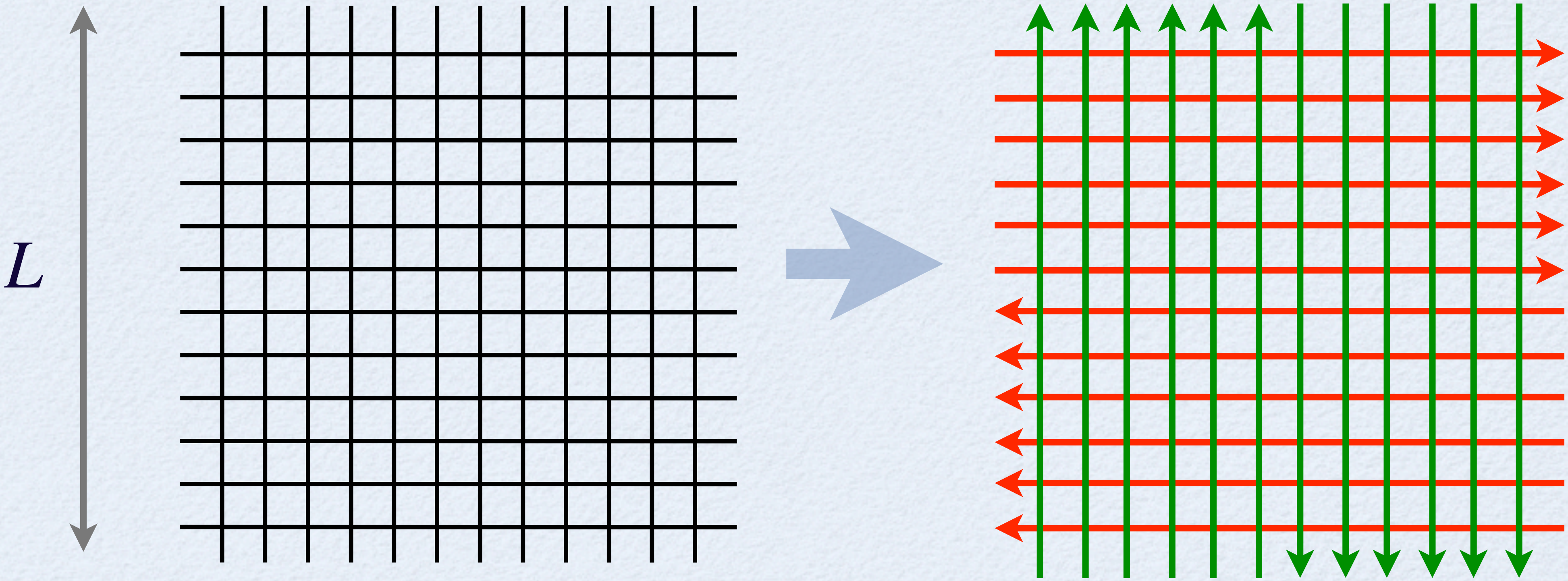
- Entropy barriers: Regions of high probability are separated by narrow paths (high probability but small entropy)



- Single region with high probability of large size (entropy)



Acceleration with loops



$T \sim L^2$

$T \sim L$

Irreversibility can significantly accelerate random walks on regular lattices.

Other approaches ?

Naive way:

- Exponentially many loops required for real systems
- Flow amplitude can not be determined based on local information

Proposed approach:

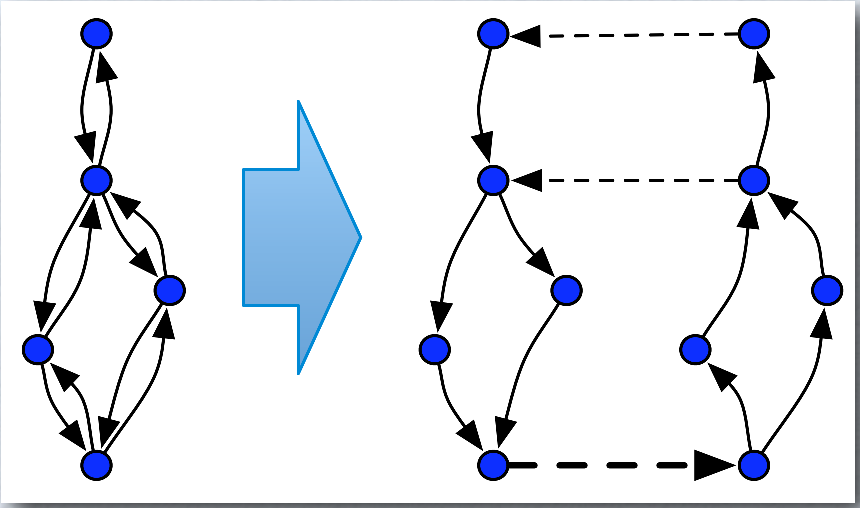
- Calculate irreversible transition probabilities “on a fly”
- Do not enforce the balance condition, instead compensate for compressibility

Skewed detailed balance

- Create two copies of the system ('+' and '-')
- Decompose transition probabilities as

$$T_{xy} = T_{xy}^{(+)} + T_{xy}^{(-)}$$

$$T_{xy}^{(+)} \pi_x = T_{yx}^{(-)} \pi_y$$



- Compensate the compressibility by introducing transition between copies:

$$\Lambda_{xx}^{(\pm,\mp)} = \max\left\{0, \sum_y T_{xy}^{(\pm)} - T_{xy}^{\mp}\right\}$$

Skewed detailed balance 2

- Extended matrix satisfies balance condition and corresponds to irreversible process:

$$\hat{\mathscr{T}} = \begin{pmatrix} \hat{T}^{(+)} & \hat{\Lambda}^{(+,-)} \\ \hat{\Lambda}^{(-,+)} & \hat{T}^{(-)} \end{pmatrix}$$

- Random walk becomes non-Markovian in original space.
- System copy index is analogous to momentum in physics: diffusive motion turns into ballistic/super-diffusive.
- No complexity overhead for Glauber and other local dynamics.

Spin cluster model

Simple example: classical Ising model defined on a full graph:

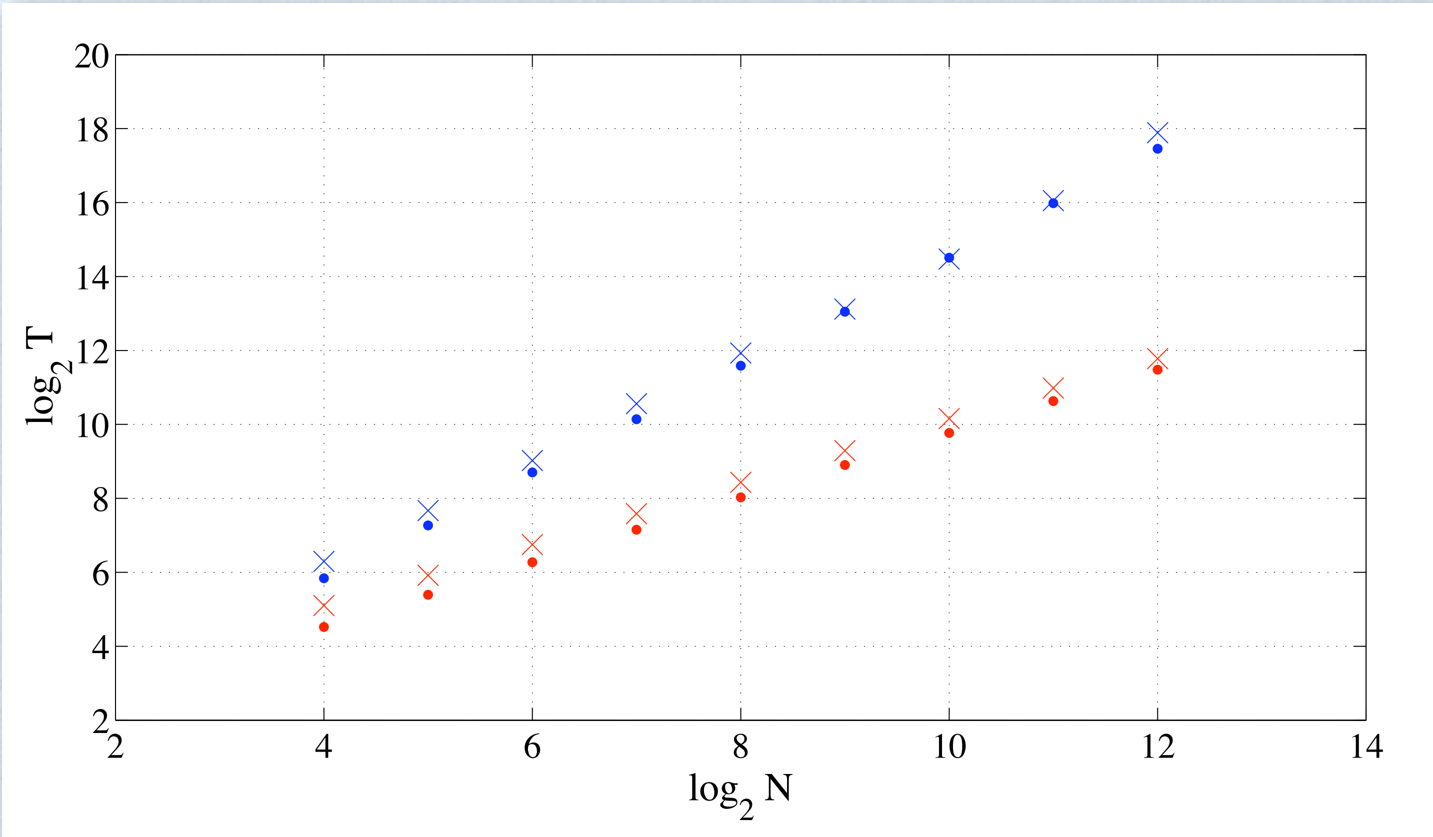
$$\pi_{s_{\square}..s_N} = Z^{-\square} \left[\frac{J}{N} \sum_{i,j} s_i s_j \right]$$

System experiences phase transition at $J=1$. Anomalous fluctuations of magnetization $\delta S \sim N^{3/4}$ lead to critical slowdown of Glauber dynamics: $T \sim N^{3/2}$

Irreversible dynamics: flip only positive spins in first copy, and only negative in second.

Spin cluster model: results

Dynamics is strongly accelerated: convergence time (defined via correlation function of *S*) decreases to $T \sim N^{3/4}$ ($T \sim N^{0.85}$ in simulations)

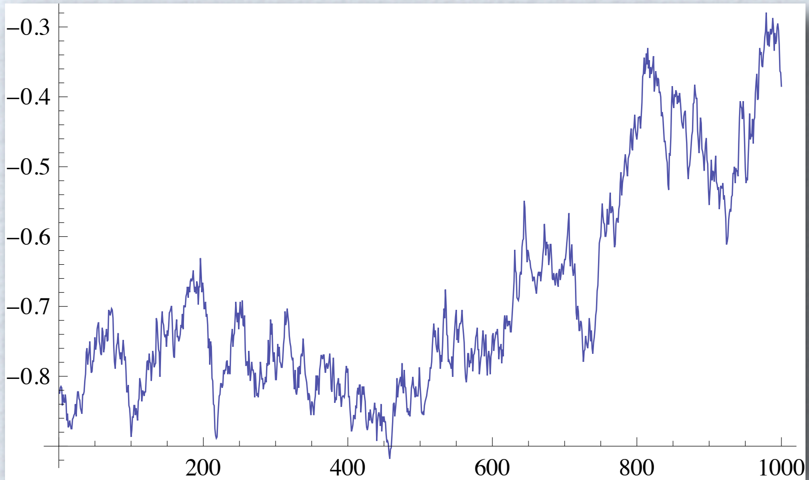


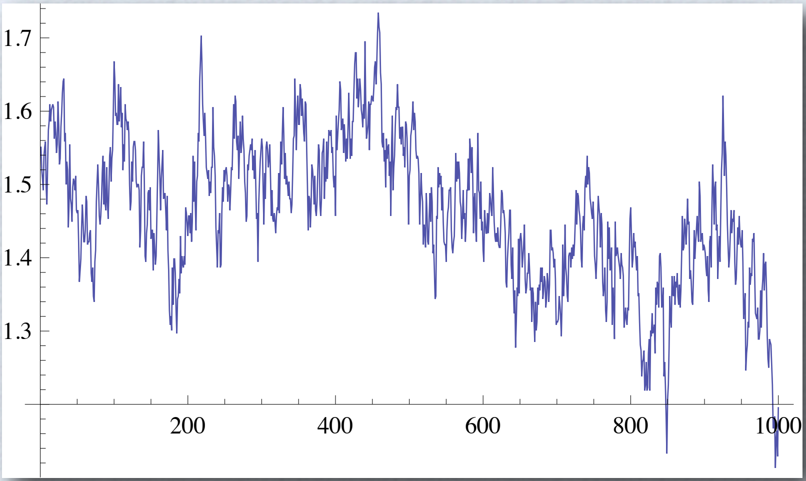
Ising model

Two-dimensional Ising model shares a lot of properties in the critical point. One can try the same algorithm.

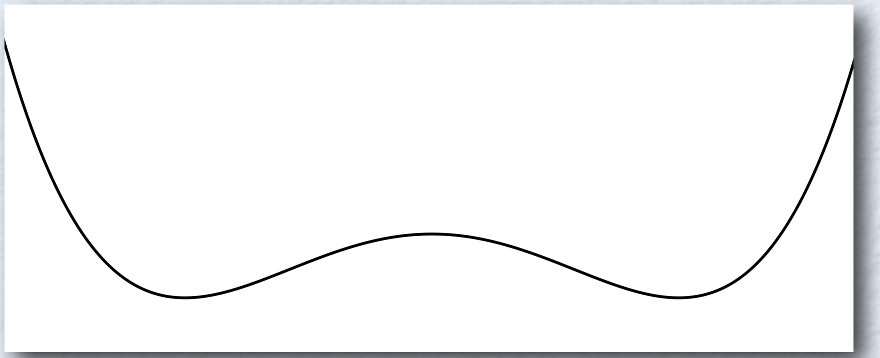
Constant factor acceleration is observed (~*3x*), however the constant does not depend on system size.

Flipping between the copies happens too frequently ($T \sim L^{1/2}$)





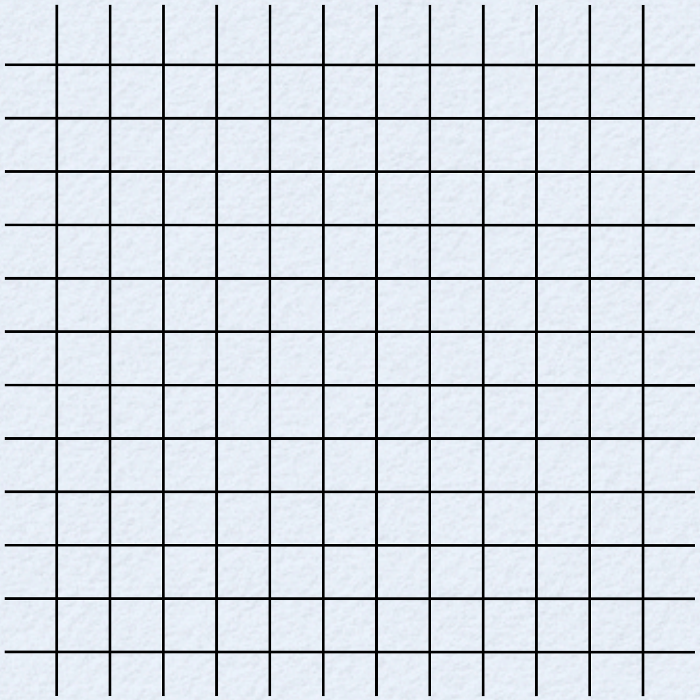
Extensions 1

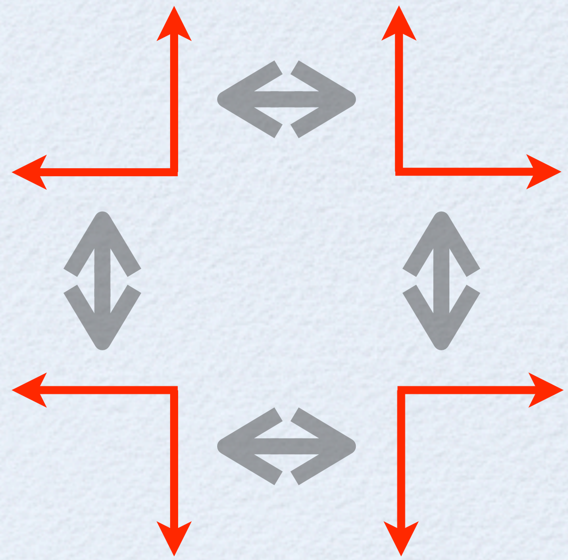


Mix irreversible fluxes in the same direction (2N copies instead of 2)

$$\sum_y \left[T_{yx}^{(\alpha)} \pi_x - T_{xy}^{(\alpha)} \pi_y \right] - \sum_\beta \Lambda_{xx}^{(\alpha,\beta)} \pi_x = 0$$

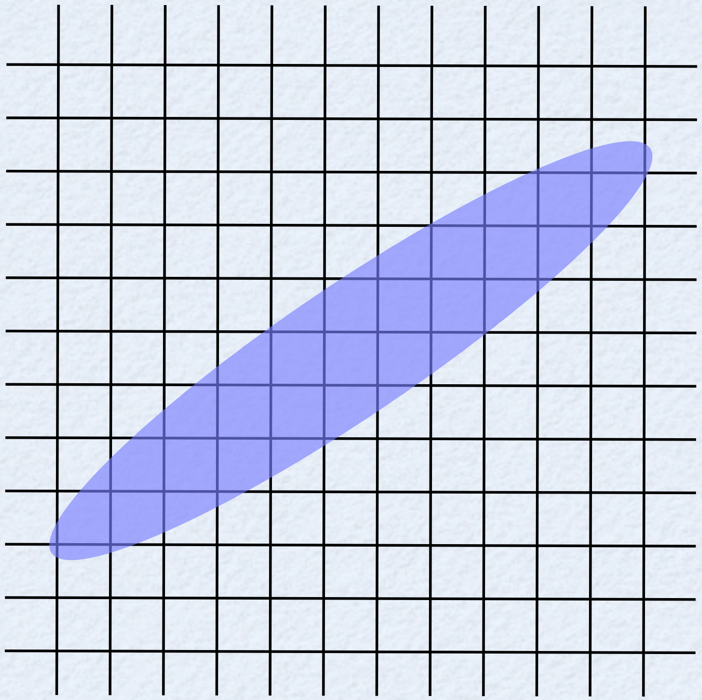
Extensions 2

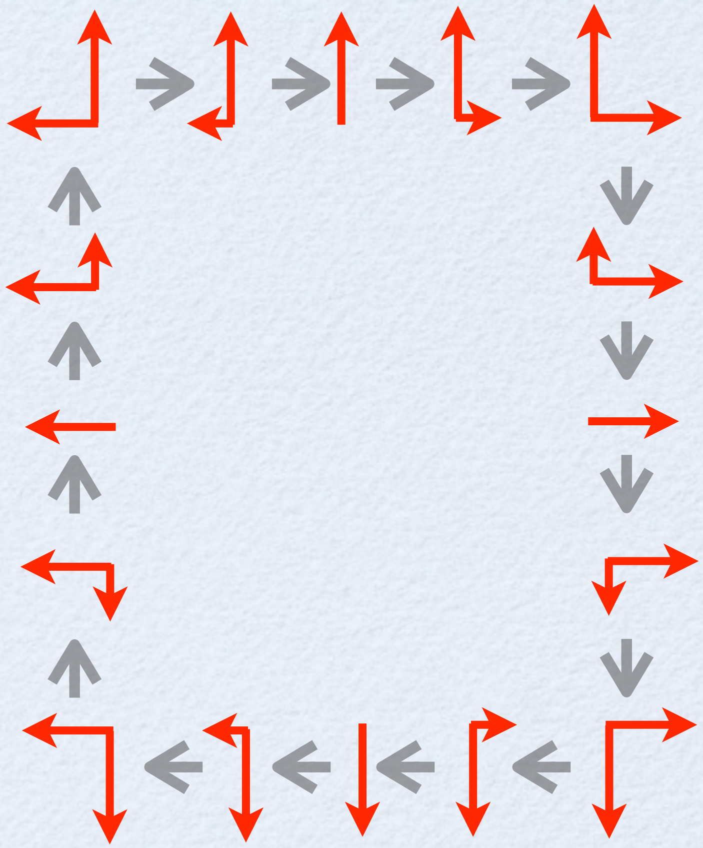




Mix irreversible fluxes in different directions (i.e. horizontal and vertical).

Extensions 3





Introduce generalized "momentum" variable.
Break symmetry in momentum space.

Phase transitions

- Explore the full phase space high-wavelength spatial harmonics of the order parameter (i.e. magnetization)
- Make the dynamics more adaptive:

Reversible dynamics : $\partial_t M_k = -\Gamma_k M_k + \xi_k(t)$

$$\Gamma_k \sim k^\chi, \quad k \to 0$$

Irreversible (Hamiltonian): $\partial_{tt} M_k = -\Gamma_k^2 M_k$

- Separate momentum variables for different parts of space

Other approaches ?

Broken DB:

- Lifting operation (*Chen, Lovasz, Pak 99*) - theoretical limits of acceleration ($T_{irr} > \sqrt{T_{rev}}$). Some toy models: (*Diaconis, Holmes, Neal 97*) .Applications to distributed computing: (*Jung, Shah, Shin 07)*
- Hamiltonian (Hybrid) Monte Carlo (*Horvath, Kennedy 88*) - continuum limit of our construction.
- Successive over-relaxation (*Adler 81*), sequential updating (*Ren, Orkoulas 06*) - another way of producing irreversible fluxes.

Reversible algorithms:

- Cluster algorithms (*Swendsen, Wang 87*) - teleportation instead of ballistic motion
- Simulated tempering (*Marinari, Parisi 92*) - several copies of the system, but with different distributions